# FAKE NEWS DETECTION USING PYTHON

## TEAM MEMBER

## 950621104073:S.Rathika

## 

## Phase 2 Submission Document

# PROJECT:FAKE NEWS DETECTION



# Content for Project Phase 2:

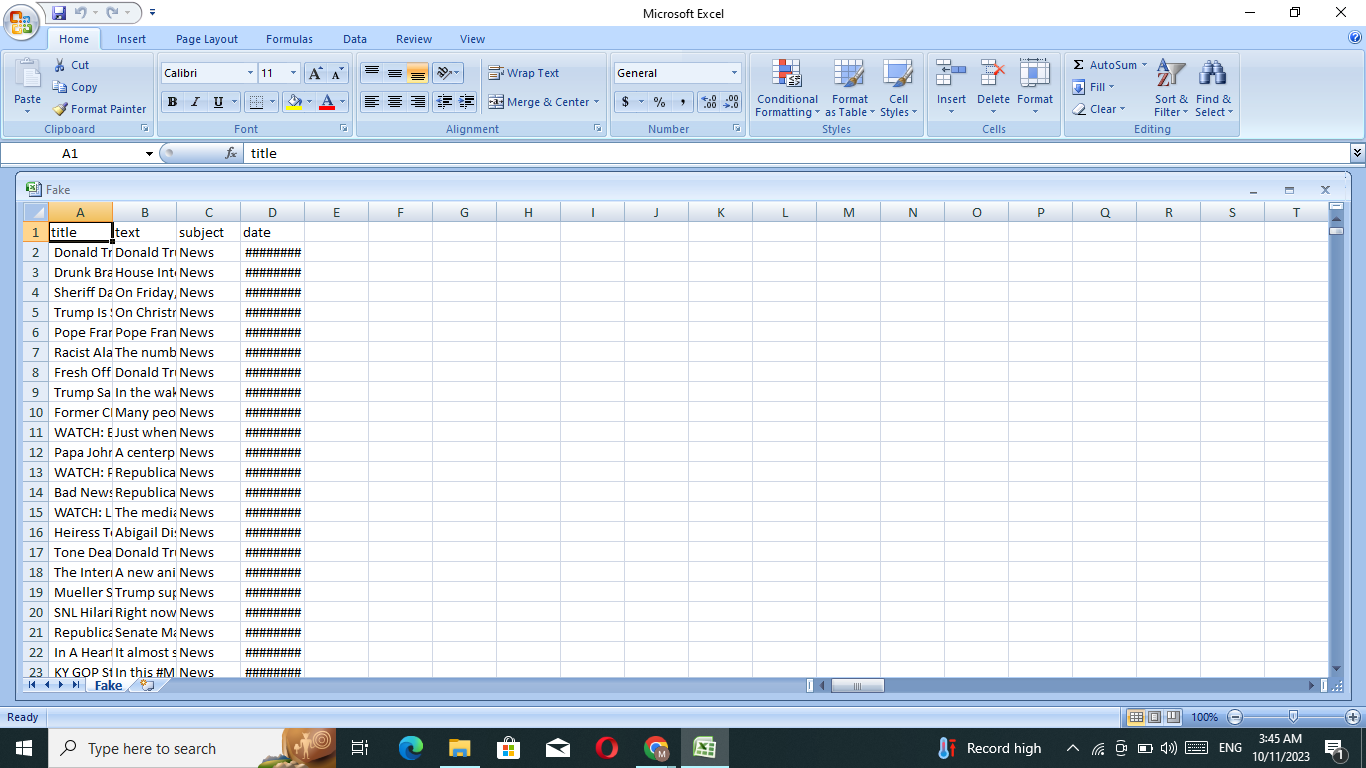
.

# Data Source :

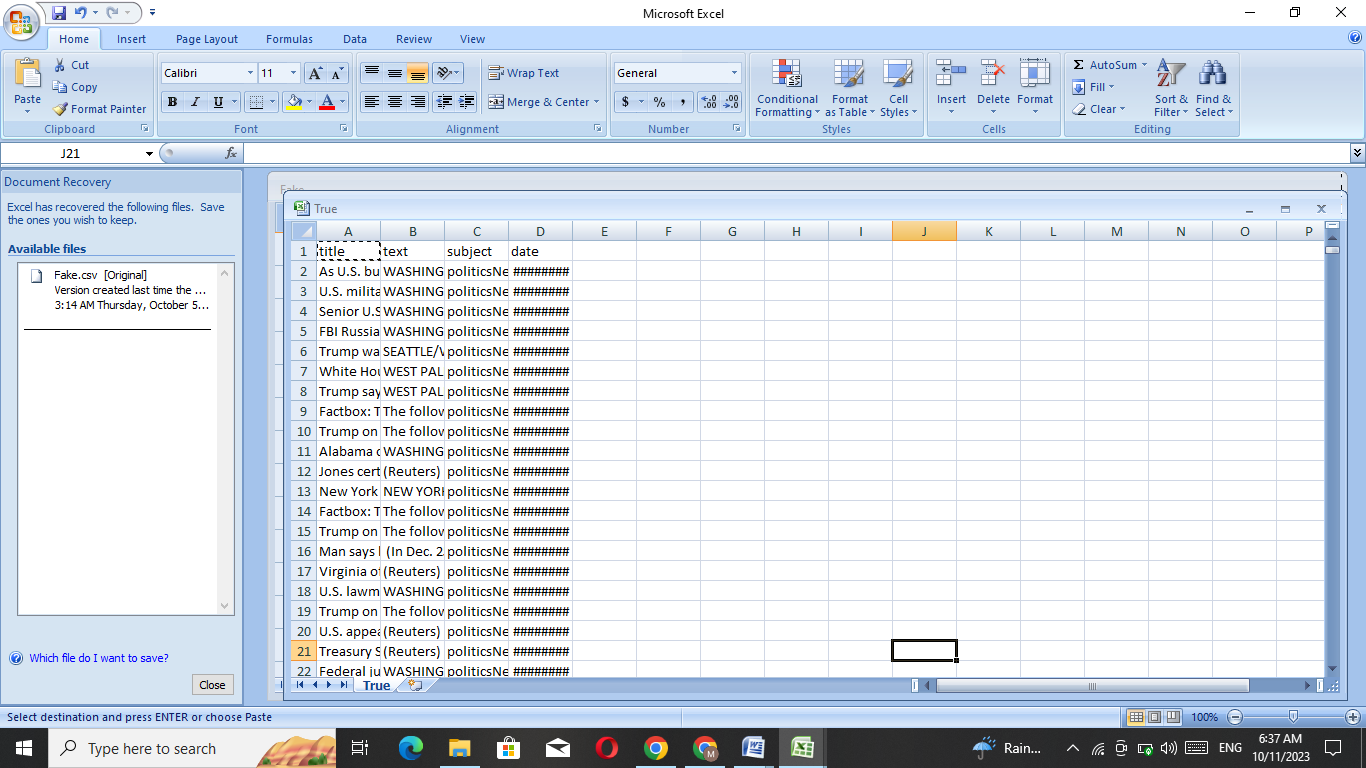
## A good data source for fake news detection using machine learning should be accurate time, location, depth,and accurancy

## Dataset Link: (<https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset>)

# FAKE.CSV

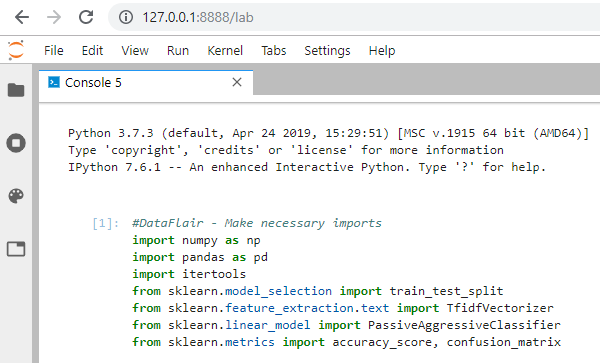


# TRUE.CSV



## INTRODUCTION:

1. Importing Libraries and Datasets
2. Data Preprocessing
3. Preprocessing and analysis of News column
4. Converting text into Vectors
5. Model training, Evaluation, and Prediction



# DEFINE:

# Fake news on different platforms is spreading widely and is a matter of serious concern, as it causes social wars and permanent breakage of the bonds established among people. A lot of research is already going on focused on the classification of fake news.

# Here we will try to solve this issue with the help of machine learning in python

# Data Collection and Preprocessing

## The first step is to collect a large dataset of news articles, both real and fake, and preprocess them by removing stop words,punctuation, and any other irrelevant information. This will helpthe AI model focus on the relevant features that distinguish real from fake news.

# Feature Extraction

## Next, various features are extracted from each article, such as sentence length, word frequency, sentiment analysis, and topic modelling. These features can be used to train machine learning models to differentiate between real and fake news.

# Training Machine Learning Models

## Once the relevant features have been extracted, they are fed into machine ]learning algorithms, such as Random Forest, Support

## Model Evaluation

## After training the machine learning models, their performance is evaluated using metrics such as accuracy, precision, recall, and F1-score. The models are fine-tuned based on the results to improve their performance.

# Deployment

## Finally, the trained models are deployed in a production environment where they can automatically analyse new news articles and classify them as real or fake. This can be done through web scraping tools or by integrating the models with existing news aggregator platforms

## PROGRAM:TO CREATE A FAKE NEWS DETECTION USING PYTHON

## FAKE NEWS DETECTION USING PYTHON

## import numpy as np

## import pandas as pd

## import itertools

## from sklearn.model\_selection import train\_test\_split

## from sklearn.feature\_extraction.text import TfidfVectorizer

## from sklearn.linear\_model import PassiveAggressiveClassifier

## from sklearn.metrics import accuracy\_score, confusion\_matrix

# data loction:

## df=pd.read\_csv(‘D:\fake news detection’)

## The process of data cleaning and data analysis

## df.shape

## df.head()

## labels=df.label

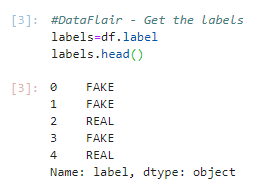
## labels.head()

## Data Cleaning

## In [3]:

## df.duplicated()

## Out [3]:



## #DataFlair - Split the dataset

## x\_train,x\_test,y\_train,y\_test=train\_test\_split(df['text'], labels, test\_size=0.2, random\_state=7)

Python data science projects - split data sets

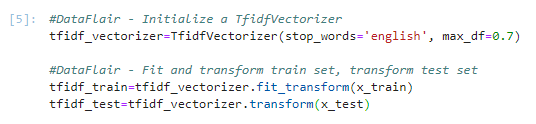
## #DataFlair - Initialize a TfidfVectorizer

## tfidf\_vectorizer=TfidfVectorizer(stop\_words='english', max\_df=0.7)

## #DataFlair - Fit and transform train set, transform test set

## tfidf\_train=tfidf\_vectorizer.fit\_transform(x\_train)

## tfidf\_test=tfidf\_vectorizer.transform(x\_test)



## DataFlair - Initialize a TfidfVectorizer

## tfidf\_vectorizer=TfidfVectorizer(stop\_words='english', max\_df=0.7)

## #DataFlair - Fit and transform train set, transform test set

## tfidf\_train=tfidf\_vectorizer.fit\_transform(x\_train)

## tfidf\_test=tfidf\_vectorizer.transform(x\_test)

## pac=PassiveAggressiveClassifier(max\_iter=50)

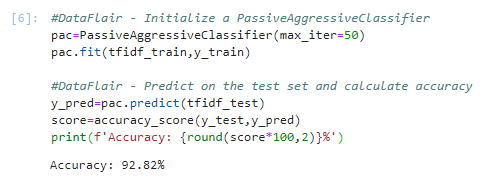
## pac.fit(tfidf\_train,y\_train)

## #DataFlair - Predict on the test set and calculate accuracy

## y\_pred=pac.predict(tfidf\_test)

## score=accuracy\_score(y\_test,y\_pred)

## print(f'Accuracy: {round(score\*100,2)}%')



## #DataFlair - Build confusion matrix

## confusion\_matrix(y\_test,y\_pred, labels=['FAKE','REAL'])

## python projects - confusion matrix

# CONCULSION:

# Overall, AI has the potential to significantly improve

# fake news detection by automating the process of identifying and

# flagging fake news articles.

